DOCUMENT RESUME

ED 294 894 TM 011 525

AUTHOR Cunningham, Tandra Tyler
TITLE Student Achievement and Teaching Performance Ratings Using a State-Wide Teaching Rating Scale: A Case of Science Teaching in the Elementary School.
PUB DATE Apr 88
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Elementary School Students; Grade 3; Grade 4; Grade 5; *Science Teachers; *Student Teachers; *Teacher Evaluation; Undergraduate Students
IDENTIFIERS Iowa Tests of Basic Skills; *North Carolina Teaching Performance Appraisal

ABSTRACT

A study involving 14 regular classroom student science teachers assigned to Grades 3, 4, and 5 and the public school students in each of the classes was undertaken to assess the relationship between student achievement and teacher performance ratings. The student teachers were all female and were seniors in a four-year undergraduate teacher education program. The classes taught by these student teachers ranged in size from 24 to 35 students. The performance ratings involved administration of the North Carolina Teaching Performance Appraisal Instrument (a statewide teacher rating scale). While teachers taught on eight consecutive school days, pairs of trained observers rated their performance. The day following the eight lessons, a 46-item objectively scored achievement test covering the content of the lessons was administered to the students in each of the 14 classes. Student scores on the Iowa Tests of Basic Skills (ITBS) were also used. Results indicate that the correlation between the total battery score on the ITBS and student achievement on the classroom achievement was not statistically significant and that the potency of teachers' skills in producing large effect sizes in the achievement of their students may be greater than previously thought. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294894

Student Achievement and Teaching Performance Ratings Using a State-Wide Teaching Rating Scale: A Case of Science Teaching in the Elementary School

Paper Presented At

American Educational Research Association

New Orleans, Louisiana

Tandra Tyler Cunningham

West Georgia College

April 7, 1988

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TANDRA TYLER CUNNINGHAM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 011 525

Student Achievement and Teaching Performance Ratings
Using a State-Wide Teaching Rating Scale:
A Case of Science Teaching in the Elementary School

Introduction

Claims for the validity of ratings obtained from teaching performance rating scales have been lodged largely in the hoped for correlation between the ratings received by teachers and the achievement of their students. Meaningful data to test this relationship between teaching performance ratings and student achievement, particularly with those teaching performance rating scales that have been adopted for state-wide use, have been difficult to come by for two main reasons.

First, tests of student achievement that are sensitive to the lessons teachers teach have not been generally available for sizeable groups of students or classrooms. Researchers have had to depend largely upon standardized achievement tests that generally are not sensitive to lessons taught by teachers and that are frequently questionable as to their sensitivity to the overall curriculum taught by teachers. Thus, using standardized achievement tests as measures of student achievement would lead one to expect, at best, weak associations between ratings of teaching performance and student achievement.

Second, particularly with those rating scales designed for state-wide use, ratings of the teacher's teaching performance generally have not been obtained while the teacher is teaching the specific learning tasks that the student achievement test measures.

This problem is especially acute when standardized achievement tests are used and when teaching performance is rated once or twice during the entire school year.

These problems in securing data to conduct critical studies of the validity of state-wide teaching performance rating scales have been difficult to overcome because of the difficulty in finding situations where a sizable number of teachers are teaching the same lessons using the same content designed to accomplish the same objectives during which they can be observed and rated on their teaching performance. The study reported here overcomes the two major flaws in previous research relating ratings on state-wide teacher performance rating scales to student achievement. Although the study reported here uses a small number of teachers, each teacher taught the same series of lessons using the same content designed to accomplish the same objectives. Each teacher's class took the same achievement test at the conclusion of the lessons--a test constructed and validated for the content taught and the objectives specified. Finally, each teacher was observed and rated on her teaching performance by trained, disinterested raters during the teaching of the lessons.

Method

Subjects. Fourteen regular classroom student teachers assigned to grades 3, 4, and 5 and the public school students in each of their classes participated in this study. The student teachers were all female and were all seniors in a four year undergraduate teacher education program. The classes taught by these student teachers ranged in size from 24 to 35 students with a median class size of 29.5. The median grade equivalent achievement scores on the Iowa

Test of Basic Skills (Total) for the 14 classes participating in this study when tested in the spring of the year when the study was conducted were 3.6 (tested at 3.7) for third graders, 4.7 (tested at 4.7) for fourth graders, and 5.6 (tested at 5.6) for the fifth grade class. The student teachers were assigned to classrooms at random.

Procedure. This study took place in the spring. After taking over full teaching responsibilities for their class, each student teacher taught, on eight consecutive school days, eight 50-minute science lessons on simple machines. The students in these classes had not studied this content previously. A previous research study established that the content and objectives of these lessons were at the appropriate level for students at these grade levels (Wilkerson, 1986). Although the content and objectives for each of the lessons were prescribed, each teacher was allowed to teach the lessons as she judged best.

While these prescribed science lessons were being taught, pairs of trained observers came unannounced to each teacher's classroom to observe and rate the teacher's teaching performance. The study reported here is part of a larger study involving student teaching during which each of these student teachers was observed and rated three times--once during the first week of student teaching prior to the teaching of these prescribed science lessons, once during the teaching of these science lessons, and once during the last week of student teaching after teaching these science lessons (Cunningham, 1987).

The day after completing the 8 lessons a 46-item objectively scored achievement test covering the content of these lessons was given to the students in each of the 14 classes.

Measuring Instruments. Ratings of teaching performance were obtained using the North Carolina Teaching Performance Appraisal Instrument (TPAI) (White, et.al, 1983). This rating scale, derived from a synthesis of the process-product research on teaching, is designed for state-wide use. The TPAI uses a five point scale to rate the teaching performance of beginning teachers on five teaching functions: Management of Instructional Time, Management of Student Behavior, Instructional Presentation, Monitoring Student Performance, and Providing Instructional Feedback (see Symposium handout). The evidence for the validity of this teaching performance rating scale has been lodged primarily in its content validity. Previous research using generalizability procedures has provided estimates of the reliability of ratings using two raters to range between .77 and .87 (Wyne, et.al., 1985). To assure adequate reliability of the ratings, each teacher was rated simultaneously by two raters with the average of the two ratings being used as the measure. For this study, the median inter-rater agreement for pairs of raters across the five rated teaching functions was .90.

Student achievement on the content taught in the 8 science lessons on simple machines was measured using a 46-item objectively scored test. Previous research on the development of this test provided substantial evidence for the content validity of this test and a KR-20 reliability estimate of .85 for scores of students at these grades levels (Wilkerson, 1986).

Results

Table 1 presents the means, standard deviations, and the minimum and maximum scores on each of the measures used in this study. The values reported for the classroom achievement test (ACH)

and the Iowa Test of Basic Skills (ITBS) are based on class averages for the fourteen classes participating in this study. The values reported for the ratings on the five teaching functions (F1 through F5) are based on the scores of the fourteen student teachers participating in this study. The statistics for these measures indicate that the values obtained were near optimal for detecting relationships between ratings of teaching performance and student achievement. The mean ratings obtained on each of the teaching functions were near the middle (3.0) of the five point rating scale used, and the range of ratings on each teaching function ranged from a low of 2.0 to highs of 4.0 or 5.0. The classroom achievement test also produced near optimal values with a mean of 34.7 (indicating an average class mean for the fourteen classes of 75 percent of the items answered correctly) with class means ranging from 27.3 (59 percent of the items answered correctly) to 39.8 (87 percent of the items answered correctly).

The correlations between the ratings on the five teaching functions (F1 through F5) and student achievement on the classroom test (ACH) on the science lessons taught may be observed in Table 2. These correlations ranged from .78 (F1 and ACH) to .44 (F4 and ACH). All of these correlations were statistically significant ($p = .05$, one tailed test) except the correlation between monitoring student performance (F4) and achievement (ACH). To provide some evidence that the classroom placement of these student teachers did not result in the ability of the students in their classes being related to the teaching skill of the student teachers, correlations were calculated between mean class scores on the Total Battery of the Iowa Test of Basic Skills (ITBS), given to all classes in the school

system shortly after the completion of this study, and the teaching performance ratings on the TPAI. As can be seen in Table 2, none of these correlations were statistically significant, ranging from -.24 (F3 and ITBS) to .07 (F1 and ITBS).

Finally, a multiple correlation analysis was done to ascertain the relationship between rated performance on the five teaching functions taken together and student achievement on the classroom achievement test, controlling for overall student ability. Overall student ability for each class was estimated using the mean Total Battery scores on the Iowa Test of Basic Skills (ITBS). The multiple correlation of the five rated teaching functions (F1, F2, F3, F4, and F5) and ITBS-Total Scores with class mean scores on the classroom achievement test was .87 accounting for 76 percent of the variance on the classroom achievement test. After controlling for the correlation between the ITBS-Total Battery scores and scores on the classroom achievement test (r=-.24), the variance accounted for on the classroom achievement test by the five teaching functions was .70 ($F_{5,7} = 4.03$; p .05).

Discussion

The idea that classroom achievement tests should test what the teacher teaches, that there should be a one to one correspondence between what the teacher teaches and what is on the test, is axiomatic. Yet, studies examining the relationship between rated teaching performance and student achievement has called attention to the fact that the formal tests, particularly standardized achievement tests, that we depend on to assess student achievement are not closely aligned with what is taught. Indeed, in this study, the correlation between the total battery score on the Iowa Test of

Basic Skills and student achievement on the classroom achievement tests was not statistically significant (r=.-24) (See Figure 1). This result is not particularly surprising since the student teachers who taught the lessons for which the classroom achievement test was designed taught these classes, full time, for a very brief period of time (3-5 weeks) and the content covered by the two tests--the classroom achievement test and the Iowa Test of Basic Skills--had little direct overlap.

This study, although done with a very limited sample of teachers, suggests that the potency of teachers' skill in producing large effect sizes in the achievement of their students may in fact be greater than previously thought.

References

Cunningham, T. (1987). Results of Effective Teaching Instruction on the Performance Ratings of Student Teachers in Mainstreamed Regular Classrooms. Dissertation, School of Education, University of North Carolina at Chapel Hill, Chapel Hill, NC

White, K., et.al. (1983). Teaching Effectiveness Evaluation Project: Final Report. Raleigh, NC: Quality Assurance Program of North Carolina.

Wilkerson, R. (1986). An Evaluation of the Effects of the 4MAT System of Instruction on Academic Achievement and Retention of Learning. Dissertation, School of Education, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Wyne, M.D., et.al. (1985). Carolina Teaching Performance Assessment System: Results of a Limited Field Test. Raleigh, NC: Quality Assurance Program of North Carolina.

Table 1: Means, Standard Deviations, and Minimum and Maximum Scores on Measures Used

Measure	Mean	SD	MINIMUM	MAXIMUM
Management of Instructional Time (F1)	3.2	.96	2.0	4.5
Management of Student Behavior (F2)	3.3	.89	2.0	4.5
Instructional Presentation (F3)	3.2	.67	2.0	5.0
Monitoring Student Performance (F4)	3.0	.57	2.0	4.0
Providing Instructional Feedback (F5)	3.1	.62	2.0	4.0
Science Achievement (ACH)	34.7	3.71	27.3	39.8
Iowa Test of Basic Skills-Grade Equivalent Total (ITBS)	4.2	.64	3.2	5.2

N=14

Table 2: Correlations Between Measures

	F1	F2	F3	F4	F5	ACH	ITBS
F1	1.00	.83*	.60*	.51*	.56*	.78*	.07
F2		1.00	.42	.47*	.45	.54*	.20
F3			1.00	.22	.34	.70*	-.24
F4				1.00	.56*	.44	.12
F5					1.00	.51*	-.18
ACH						1.00	-.24
ITBS							1.00

*p=.05 (1 tail); r≥.46

F1- Rating on Management of Instructional Time
F2- Rating on Management of Student Behavior
F3- Rating on Instructional Presentation
F4- Rating on Instructional Monitoring
F5- Rating on Instructional Feedback
ACH-Class Mean Score on Science Achievement Test
ITBS-Class Mean on Grade Equivalent Score on Total Battery of Iowa Test of Basic Skills

Figure 1: Scattergram of Correlation Between Science Achievement Test Score and ITBS Grade Equivalent